

EXHIBIT 120

AWBid AdSpam/Publisher Quality Investigation

December 2014

Petter Premberg (petter@), Diogo Farinha (diogofarinha@), Stephen White (stwhite@), Adrienne Walcer (apwalcer@)

Confidential - Do not share or redistribute

Overview

AdSpam concerns

Terminated AdSense publishers

Arrow blacklisted domains

Missing signals

Auto-refreshing ads

Debiting bad publishers

Policy concerns

Content URLs pointing to ad servers

Webmail

Low quality pages

...

Overview

We have investigated the traffic and publisher quality of AWBId, both the program as a whole and the individual exchanges that have accrued traffic through AWBId. For some background see [this document](#).

The exchanges investigated were Rubicon, OpenX, PulsePoint and Pubmatic for the dates of November 9-22 2014¹. Other exchanges that have been bidded on (Casale, AdShield and YieldLab) had less than 4,000 clicks in this 14 day period, which makes it impossible to reliably assess the quality of their traffic from a spam perspective, and also may not give a good estimation of what pages they would be serving ads on if their traffic volume was to reach a normal level. For observations of the individual exchanges, see these documents:

- [Rubicon](#)
- [OpenX](#)
- [PulsePoint](#)
- [Pubmatic](#)

¹ PulsePoint was already blacklisted at this point, so for this exchange an earlier date range was investigated

In this document we'll go through some of the main concerns we have with AWBId traffic overall, and some ideas of how to deal with these once the traffic starts to ramp up. Currently we bid at different levels (changing slightly over time), but for the investigated period these were the numbers of clicks observed for each exchange:

Exchange	Clicks	%AutoSpam	%ManualSpam	Cost	Credible GradeWithCI	Credible Coverage
Rubicon	310,704	23.37%	2.3%	\$81,641.19	0.703 (0.677, 0.729)	0.29
OpenX	106,927	18.98%	3.5%	\$53,358.98	0.839 (0.797, 0.882)	0.36
Pubmatic	25,791	21.85%	52%	\$7,056.92	0.655 (0.580, 0.737)	0.31

We previously bid on PulsePoint as well, however this whole exchange was blacklisted due to low quality traffic. These are the numbers for this exchange before this happened (October 1-19):

Exchange	Clicks	%AutoSpam	%ManualSpam	Cost	Credible GradeWithCI	Credible Coverage
PulsePoint	38,934	13.65%	94%	\$1,110.47	0.024 (0.012, 0.049)	0.50

AdSpam concerns

As far as the traffic quality goes, we have not noticed a great volume of malicious traffic on these three now serving exchanges, except for Pubmatic which had some issues. In cases where invalid clicks and impressions have been observed, we have been able to identify this and separate the traffic that needs to be credited/blacklisted based on the spam signals that are available to us at the moment.

However, we have come across some situations that are not straightforward to deal with, particularly in terms of guaranteeing that advertisers will not suffer from known instances of spam that we have taken actions to get rid of in the past.

We should also note that for PulsePoint there were traffic issues so severe that we eventually had to stop bidding on the exchange altogether (95% of click and impression traffic was deemed as invalid). This highlights the need to carefully evaluate any new exchanges that we may consider bidding on in the future.

Terminated AdSense publishers

When disabling an AdSense account, we typically do not allow them back into the network without them successfully going through an appeals process. This is an expectation we set with AdWords advertisers as well, when they have reported some sites to us and we take action on these sites. While AWBId is not technically a Google ad network, but rather a way to bid on other exchanges, it is possible that AdWords advertisers that opt in to AWBId will be disappointed to see their ads shown on pages that they have reported to us and were told that we have disabled all bad publishers to avoid their ads from showing there in the future.

Commented [1]: PulsePoint?

Commented [2]: I excluded them as they were already blacklisted at this time period. I wanted to focus on recent traffic here as we are looking at the state of AWBId as it stands now. We could add a separate table though to show how a bad exchange might look.

Commented [3]: PulsePoint?

Commented [4]: I excluded them as they were already blacklisted at this time period. I wanted to focus on recent traffic here as we are looking at the state of AWBId as it stands now. We could add a separate table though to show how a bad exchange might look.

Commented [5]: I would also mention that we lost some signals on query pipeline: QueryHO, HistoryLength, ...

Commented [6]: Right, this is one of the things I wanted to include here but had forgotten, thanks! :) Added as it's own point below.

Commented [7]: I would mention that we investigated PulsePoint and excluded it from AWBId since the traffic was ~95% spam.

This show evidence that adspam is owning the traffic quality process and blocking perpetrator from AWBId.

Commented [8]: Very good point. It also shows the need to evaluate new exchanges carefully if we want to add some in the future. Added a few sentences here.

One way around this could be to avoid showing ads (pre-bid filtration) on pages that belong to terminated AdSense publishers, by writing a script that gets this information and creates blacklists automatically based on this. Unfortunately, we currently don't have a defined way for identifying such pages at the moment, and any way to do so automatically might lead to false positives (that is, we would avoid serving ads on pages we think might get spammy clicks and impressions based on old investigations, even though the owner of the page might have changed, spammy traffic might have stopped, etc.).

Arrow blacklisted domains

In the last couple of years, the AdSpam and Publisher Quality team has implemented several procedures to protect advertisers from abuse of arrow ad formats, ie. where users have been misled to click on an arrow ad thinking it will navigate them to a different page under the publisher's website (eg. "Next" buttons on gallery pages). The current approach is aggregating on **publisher code X ad block key** for each mobile browser class (low/high end mobile phone/tablet/desktop), so that ad blocks belonging to a publisher account that gets many clicks on arrow ads, and also performing badly, get blacklisted which means the arrow ad format is taken away and they will instead serve normal text ads.

In AWWid, publisher codes do not correspond to an individual publisher account but rather a whole ad exchange, so the distributions on pages where ads are shown under these publisher accounts are much wider than for AdSense publishers. This may cause arrow blacklists to be less effective on AWWid, in case we do not get enough click data on each ad block key to determine if it is performing well enough. One solution could be to develop a similar blacklist based on domain instead of ad block key, however this is not implemented and may be a less than ideal solution, as spammers can easily change domains if they realize that this would help them get arrow ads back.

Missing signals

Compared to AdX/AdSense, we are missing some significant signals and entities that we would use to isolate and evaluate traffic. This is particularly the case on the query side, which may make it difficult to debit for impressions based on properties from clients' machines in the future (as opposed to domain/bidder_id/etc.).

Auto-refreshing ads

On AdX, which is used by many AdWords advertisers to bid on display based networks, publishers are allowed to issue new ad requests automatically without the user manually refreshing the page only after 300 seconds. On AWWid this is not enforced at the moment, and in investigations we have noticed several domains where ads are auto-refreshed as often as every 30 seconds.

This might become an issue for advertisers who are used to AdX as the user will be less likely to actually see the ad, but it could also lead to Google taking a loss in cases where ads are served on auto-refreshing pages with much lower CTR than expected. Google pays publishers for impressions, and this cost is converted into a predicted click cost, which is meant to translate accurately into the cost we charge advertisers. Ideally this conversion should be accurate so

Commented [9]: +alexjacobson@google.com We are putting together a little list of our AWWid concerns so far, one of them is about arrow ads. Can this level of detail be shared with the product team? Or should it be changed to something less specific? Thanks!

Commented [10]: Good to share with product.

Technically its aggregated as publisher code x ad block key x mobile browser class.

Commented [11]: I'm confused why we get this query time signal at all. It's in the XBFE->Cat2 query (cafe.partnerlog) so it must be sent as part of the bid request?

Furthermore the ad block keys observed on AWWid don't seem to match non-AWWid traffic (at least the top 50 have no overlap).

I'll ask Joan today if she knows how adk= is set, perhaps it's supposed to mean something else and just coincidentally aligns with our ad_block_key query parameter.

Commented [12]: OK, Joan answered that it's this: go/awbid-adkey. Basically it's a hash of information like publisher, ad slot dimensions. They added it to help pCTR models.

It's generated server side in this case, so it's not really the same semantically as adk from our ad tags which actually looks at the DOM structure of the ad on the page.

Overall, there's no analogous signal to adk for AWWid that's available at serving time, so arrow whitelisting/modifications are not really possible without significant changes.

Commented [13]: Ouch. Thanks Sam, will update this. One question though, do you know why some AWWid clicks are being filtered by SFRC_NESSIE_WP_ABK_MBC_BLACKLIST? Is that because the blacklist is incorrectly treating the AWWid specific version of ABK as a Spam ABK, thus possibly under- or overfiltering? +alexjacobson@google.com So it seems there's some more trouble with Nessie protection on AWWid. Would you see it as problematic to change to a different entity to blacklist by, like domain instead of ABK? If so, should we turn arrows off for AWWid traffic? Currently over 20% of AWWid clicks are on arrow ads, so it might be a struggle to determine where such ads should be shown if we do

Commented [14]: The ad_block_key is filled in so the filter looks at it. It's not totally unrelated to what traditional ad_block_key tries to measure since this both are trying to identify a particular source of inventory. This one is just (probably) less fine grained.

Hard to say whether it's over or under filtering. I would say that the major attack vector of trying to disguise the arrow as a navigational element would not be as

that impressions of ads/ad slots that have low CTR will get cheaper than those with higher CTR, but when CTR is artificially impacted by something like auto-refresh in combination with the setup of AWBId, where most publisher properties (exchanges) have relatively few clicks/impressions on the same domains/ad slots.

Debiting bad publishers

So far, we have verified that we are able to credit advertisers for invalid clicks coming in through. However, this only ensures that we give money back to advertisers, while Google still eats the cost on the impression side (unless we debit for impressions as well). In some cases we do not have query information for the entities used to credit for clicks, which would make debits impossible in such cases. If we credit by domain, bidder_id or so, however, we should be able to identify this entity on the query side as well. A proposed process for this has can be found in the [AdSpam support for AWBId](#) document (see section "Refunds Process"), but this has not been tested yet.

Traffic spikes for advertisers

This only applies if AWBId was to launch as opt-out, and thus lead to a possibility that AdWords advertisers would see sharp increases in traffic without changing anything in their campaigns. This could have a negative impact on the advertisers' experience, particularly in combination with the other concerns mentioned above, regardless if the traffic is clean from a spam perspective or not. If opt-in, we could warn advertisers when turning AWBId on that this might have a significant effect on their traffic volume as they will show ads on additional inventory.

Policy concerns (see master document under construction)

Content URLs pointing to ad servers

(Not seeing where the ad was shown)

Commented [15]: +stwhite@google.com Do you have any info of how significant this is? If I understood Vegard right yesterday I think he said that overall something like 20% of the traffic on some of these exchanges came from hidden ad serving URLs. Does this seem to be the case for AWBId?

Webmail

Low quality pages

Risky for very sensitive/"brand aware" advertisers

tl:dr: PII is an important issue to resolve ASAP, we continue to see challenges with some content that we wouldn't allow on AdSense Publishers.

1) Tripwire+ Entries 11/24/2014

We're getting 2896 domains flagged for passing PII.

Publisher ID	Exchange	Status	Detected Type	# of Queries	# of Domains
ca-pub-6362111942204036	Pubmatic	DETECTED	Email,Password	731641	691
ca-pub-2208106016887066	Yieldlab	DETECTED	Email,Password	4547	62
ca-pub-3567512973010615	Adscale	DETECTED	Email,Password	243730	279
ca-pub-6579838053286784	Rubicon	DETECTED	Email,Password	103639378	597
ca-pub-5722610347565274	Casale	DETECTED	Email,Password	2565149	547
ca-pub-3028258428956246	OpenX	DETECTED	Email,Password	31848927	720

2) Seabiscuit Entries (direct policy decisions):

Where we have been able to review, we see some troubling content and behavioural violations.

Violation	%age
Deceptive General	1.22%
Fight Videos	1.22%
Google Product Abuse	1.22%
Pedophilia	1.22%
SQE Webspam	1.22%
Too many large ads (i.e. CSI abuse)	1.22%
Deceptive Site Navigation	2.44%
Links to Adult Sites	2.44%
Gore General	3.66%
Adult Comment Spam	4.88%
Fetish/Sexual Aids	4.88%
Nudity/Porn	7.32%

Commented [16]: Additionally, we don't have the technical ability to blacklist all of the URLs we've come across, which leaves advertisers exposed as these items linger.

Commented [17]: +apwalcer@google.com

Hey Adrienne, Can you please give a little bit more detail on why you can't blacklist the URLs / domains?

Thanks

Commented [18]: When violations are served through direct/NPL systems (as these accounts have been), we have to action using The Boot, which (in a url) drops everything after a '?'. This would frequently blacklist a lot of nonviolating content, which we have to avoid.

Commented [19]: +jnickerson@google.com

Actually, we were discussing this recently and the hypotheses is that The Boot would work.... John have you checked this?

Also, what if you blacklist the full domains and not just the URL?

Commented [20]: I'm discussing with publisher policy infra, and apparently the ability to blacklist at the property level won't work (yet). The tool sets the blacklist to global (all properties). More details to come...

Commented [21]: URL-level blacklisting with The Boot isn't scalable at all, which is likely why it wouldn't work, but it's how we've been dealing with these items for a few months.

Commented [22]: I might be missing something, but can we still block on domain level? I wouldn't mind being a bit extra cautious and blacklist a whole domain if some of the sites on it are violating policy/getting spammy traffic

Commented [23]: I wouldn't mind either!! But it's not really my choice to make.

Commented [24]: I think we might be able to decide on these things ourselves, as long as it makes sense from an abuse prevention perspective. Of course it might be different in some cases, if a huge website has just one policy breaching URL and others have good quality content we wouldn't want to blacklist the whole domain (not sure how often that happens), but I don't think any external parties will have a say in how we decide to defend AWBId. +diogofarinha@google.com does that sound right, based on your recent discussions with the product team?

Too Many Ads	8.54%
Unnatural Attention	18.29%
Illegal File Sharing	19.51%
Sexually Gratifying	19.51%

3) ENFDB (online policy decisions):

684 blacklistings, where review occurred and was possible (e.g. no obfuscation), for the following violations:

- Google Product Abuse
- Deceptive Site Navigation
- Violations in Dynamic Content
- Adult Links
- File Sharing
- Unnatural Attention

...